# GRAMMATICAL FORMALISM AND COMPUTATIONAL ANALYSIS OF NOMINAL COMPOUNDS IN TAMIL

Dr.S.Rajendran
rajushush@gmail.com

## INTRODUCTION

A morphologically complex word containing at least two elements which can otherwise occur as free forms (i.e. as independent words) can be considered as a prototypical compound.

e.g.

1. *talaiyaNai* (< *talai* 'head' + *aNai* 'support') 'pillow'
2. *marappeTTi* (< *maram* 'wood' + *peTTi* 'box') `wooden box'
3. *kiLippaccai* (< *kiLi* 'parrot' + *paccai* 'green') 'parrot green'

Compounding is a grammatical process by which complex words are formed from smaller elements which have word status under normal circumstances. Affixation is different from compounding as it involves morphemes which do not have word status. It is the word-like behaviour of a string of elements which indicates that it is a compound. Though rearrangement of constituents in a construction is possible in a language, the constituent parts of a compounds cannot be rearranged.

e.g.

1. *talaiyaNai* 'pillow' vs. ?*aNaitalai*
2. *marappeTTi* 'wooden box' vs. ?*peTTimaram*
3. *kiLippaccai* 'parrot-green' vs. *paccaikkiLi* 'parrot'

("?" marks indicates that the expression does not mean or refer anything.)

In addition, the constituents of a compound do not allow themselves to be separated by intervening material showing word like quality. In languages like English the compounds are distinguished from phrases by their typical stress pattern. In some languages there is linking morpheme compounding the constituents (as found in the compound *morph-o-sytnax*).

The compounds can be studied at least from five points of view:

1. based on the grammatical categories of words which constitute compounds
2. based on the semantic classes
3. based on the possible linking elements
4. based on the deep structure
5. based on the morphophonology.

Compounds which are nouns are called nominal compounds. Traditional grammarians of Tamil look at nominal compound formation in Tamil from two points of view:

1. Compound nouns are derived from phrases by the deletion of elements like case suffixes, comparative particles, tense suffixes, co-ordinate particles, and predicative elements.
2. Compound nouns are derived only by the juxtaposition of words and only for the interpretation of meaning they have to be expanded with the help of elements mentioned above.

The modern approaches, whether they represent lexicalist approach or generative approach, are not totally different from the traditional view points on nominal compounds.

## CHOOSING OF A GRAMATICAL FORMALISM

### Distinction between the root compounds and synthetic compounds

Distinction has been made in the literature between primary (root) compounds and synthetic (verbal) compounds. Primary compounds are simply concatenated words.

e.g. *kaaTu* 'forest' + *vilanku* 'animal' = *kaaTTuvilanku* 'wild-animal'

Synthetic compounds are formed from deverbal heads and non-heads which fulfil the function of the argument of the verb from which the head is derived.

e.g *vaNTi* 'cart'+ *ooTTi* 'driver'= *vaNTiyooTTi* 'cart driver'

(*ooTTu* 'drive' + *i* = *ooTTi* `driver'; -*i* is an agentive suffix.)

It has been claimed in the literature that significant difference lies between root compounds and synthetic compounds. The difficulty lies in the decision of finding what constitute a synthetic compound. There are at least two main issues in the interpretation of the synthetic compounds. The first view is that synthetic compounds are not different from root compounds and that they are formed by the juxtaposition of words out of which the head word is one which is derived from a verb. But it has been argued that taking this position will deny or suppress the relation between the deverbal head and its arguments. For example, the relation between the following two phrases will be suppressed if we consider the synthetic compound at par with the root-compounds.

e.g.

1. *vaNTi ooTTi* `cat driver'
2. *vaNTi ooTTu* `drive a cart'

The second view is that the synthetic compounds are different from the root compounds as they are derived from an underling phrases or sentences in which the relation between the head element, i.e. verb, and the non-head element, i.e. noun, is clearly shown. In the second view also two kinds of stand have been taken by the scholars: one group take up the position of lexicalist and the other group take up the position of generativist. The lexicalist position is that the synthetic nominal compounds are derived by morphological process of word-formation and the generative position is that the compounds are derived by syntactic process.

Morphological approach is taken up by Roper and Siegel (1978), Selkirk (1982), Lieber (1983) and Di Sciullo and Williams (1987) and syntactic approach is taken up by Fabb (1984), Sproat (1985), and Roeper (1988). A detailed account of the two controversial approaches is found in Spencer (1991). As for as the computational approach is concerned what matters is the exposition of these compounds in terms of simple mathematical or grammatical formalism and the mechanism by which the meanings of the compounds can be interpreted.

The position which has been taken up here is that there is no difference between the formation of root compounds and synthetic compounds and both are formed by the rule of the following:

$$N + N \longrightarrow N$$

The fact that the second head noun is derived from a verb will be taken care of by the word-formation of rule of the following type:

$$V + \text{Suffix} \longrightarrow N$$

A unification approach is taken up here in which the syntactic as well as semantic features associated with the verb will be unified with the non-head noun and thus the properties of the verb will be carried over to the compound also; the acquired idiosyncratic properties will be taken care of in the process of lexical formation and the interpretation. For example, the information that *oTTu* 'drive' is in objective/thematic relation with the noun *vaNTi* 'cart' will be available for the derived noun *ooTTi* also.

We are not going to take up the transformational generative grammarians' approach of deriving compounds from underlying deep structures which carry all the properties necessary for the semantic interpretation of the compounds. The generation or derivation of the nominal compounds will be done by the simple rule of juxtaposing nouns. The weeding out of uninterpretable or unacceptable compounds will be taken care of at the level of semantic interpretation which is known in the TG frame work as LF (i.e. Logical Form).

**Traditional outlook**

The traditional grammarians of Tamil classify the compound nouns into six types based on the semantic interpretation or the recoverability from the presumed sources: *veeRRumaittokai* 'casal compound', *uvamaittokai* 'compound of comparison', *vinaiyin tokai* 'verbal compound', *paNpin tokai* 'compound of quality', *ummaittokai* 'compound involving *um*' or co-ordinate compound and *anmozhittokai* '*bahuvriihi* compound'. According to one interpretation, nominal compounds are considered as reduced forms of phrases in which the elements like case suffix, comparative particles, tense, particles of quality, co-ordinate particles, etc. are dropped. According another interpretation, it is not that the compounds are derived from phrases, but only for interpretation of meaning they are expanded into phrases. The table given below will exemplify the above mentioned traditional view on nominal compounds:

| Sr No | Compound | Possible expansion | The type base on the deleted or recoverable element |
|---|---|---|---|
| 1. | *makrantappai* ‘pollen grains’ | *makarantattaiyuTaiya pai* ‘bag which contains pollen grains’ | Casal compound in which the second case suffix *-ai* is recovered or deleted |
| 2. | *maNal meeTu* ‘sand hill’ | *maNalaal amainta meeTu* ‘elevated place made up of sand’ | Casal Compound in which the third case suffix *-aal* is recovered or deleted |
| 3. | *makkaL kalvi* ‘education for people’ | *makkaLukku kalvi* ‘education to the people’ | Casal compound in which the fourth casal suffix *-ku* is recovered or deleted |
| 4. | *mara eNNey* ‘wood oil’ | *marattiliurntu eTukkappaTTa eNNey* ‘oil taken from wood’ | Casal compound in which the fifth case suffix *-illiruntu* is recovered or deleted |
| 5. | *mazhai aLavu* ‘amount of rain fall’ | *mazhaiyatu aLavu* ‘amount of rain fall’ | Casal compound in which the sixth case suffix *-atu* is recovered or deleted |
| 6. | *marappuzhu* ‘wood-worm’ | *marattinkaN vaazhum puzhu* ‘worm living in wood’ | Casal compound in which the seventh suffix *-kaN* is recovered or deleted |
| 7. | *cuTuniir* ‘hot water’ | *cuTTaniir/ cuTukinRa niir/ cuTum niir* ‘ water which became /becomes / will become hot’ | Verbal compound in which the past tense suffix *-T + a* or the present tense suffix *-kinR + a* or future tense suffix *-um* is recovered or deleted |
| 8. | *karunkal* ‘block of stone’ | *karumaiyaana kal* ‘black stone’ | Compound of quality in which the suffix *-mai* and the paticle *-aana* are recovered or deleted |
| 9. | *kuzhal viLakku* ‘tube-light’ | *kuzhal poonRa viLakku* `tube like lamp’ | compound of comparison in which the comparative particle *poonRa* is recovered or deleted |
| 10. | *varavu celavu* ‘expenditure’ | *varavum celavum* ‘income and expense’ | Co-ordinate compound in which the co-ordinate element *-um* is recovered or deleted |
| 11. | *koTiyiTai* ‘woman with slim waist’ | *koTi poonRa iTaiyai uTaiya peN* ‘woman with creeper like waist’ | *Bahuvriihi* or exocentric compound in which the comparative particle *poonRa*, the second case suffix *-ai*, the linking vebal element *uTaiya* `having’ and the exocentric head noun *peN* ‘woman’ are recovered or deleted |

We are not concerned with exocentric compounds here. It should be noted that in all the above mentioned type of compounds not only the suffixes and particles, but also verbal elements are recovered or deleted for the interpretation or derivation of the compound nouns. The comparative element *poonRa* `as if/like' and the quality element *aana* can be interpreted as verbal element. Even in the examples like *makkaL kalvi* `education for people' and *mazhai aLavu* `amount of rain fall' where it appears as if only the case suffixes have to be recovered or deleted, we can posit verbal element like *uriya* `belonging to'. It is clear from the above mentioned outlook of traditional grammarians that certain elements have to be recovered for the semantic interpretation of nominal compounds and the most important of them is verbal element.

It should be remembered here that the elements, especially the predicative or verbal elements, to be recovered for the interpretation of meaning can be predicted by native speakers, though the predictable variables may be numerous. A theory which generates compounds from underlying representations does allow variables in the recovery of the compounds. But the stand that the compound nouns are derived by mere juxtaposition of nouns and that the validation of the possible combinations are left to the decision of mental lexicon which interprets the meaning, can incorporate variables. But here also the native speakers intution allows only a few variables which can be interpreted as synonyms belonging to a semantic area. [Note down the similarity of this position with that of Levi (1978), Allen (1979) and Leiber (1980).]

**Levi's approach**

A brief account of Levi (1978) is worth mentioning here as the formalism we are going to adopt here is having some bearing on the account given by Levi. Levi posits a set of nine 'specifiable predicates' that are 'recoverably deletable' in the process of formation of complex nominals (abbreviated as CN). The nine 'Recoverably Deletable Predicates' (abbreviated as RDP), according to her are CAUSE, HAVE, MAKE, USE, BE, IN, FOR, FROM and ABOUT. The examples given by her are listed in the table given below for reference (Levi, 1978:76):

| Sl. No | RDP | $N_1$ < direct object of relative clause | $N_1$ < subject of relative clause |
|---|---|---|---|
| 1. | CAUSE | disease germ | birth pains |
| | | tear gas | drug deaths |
| | | malarial mosquitoes | nicotine fit |
| | | traumatic event | viral infection |
| | | mortal blow | thermal stress |

| | | | |
|---|---|---|---|
| 2. | HAVE | picture book | government land |
| | | apple cake | lemon peel |
| | | gunboat | student power |
| | | musical comedy | reptilian scales |
| | | industrial area | feminine intution |
| 3. | MAKE | honeybee | daisy chains |
| | | silkworm | snowball |
| | | musical clock | consonantal patterns |
| | | sebaceous glands | molecular chains |
| | | songbird | stellar configurations |
| 4. | USE | voice vote | |
| | | steam iron | |
| | | manual labour | |
| | | solar generator | |
| | | vehicular transportation | |
| 5. | BE | soldier ant | |
| | | target structure | |
| | | professional friends | |
| | | consonantal segment | |
| | | mammalian vertebrates | |
| 6. | IN | field mouse | |
| | | morning prayers | |
| | | marine life | |
| | | marital sex | |
| | | autumnal rains | |
| 7. | FOR | house doctor | |
| | | arms budget | |
| | | avian sanctuary | |
| | | aldermanic salaries | |
| | | nasal mist | |
| 8. | FROM | olive oil | |
| | | test-tube baby | |
| | | apple seed | |
| | | rural visitors | |
| | | solar energy | |
| 9. | ABOUT | tax law | |
| | | price war | |
| | | abortion vote | |
| | | criminal policy | |
| | | linguistic lecture | |

The possible expansion for each type of compound could be as exemplified below:

| | | |
|---|---|---|
| tear gas | = | gas which causes tear |
| drug deaths | = | deaths which are caused by drug |
| picture book | = | book which has picture |
| government land | = | land which government has |
| honeybee | = | bee which makes honey |
| daisy chains | = | chains which are made up of daisy |
| voice vote | = | vote which uses voice |
| soldier ant | = | ant which is a soldier |
| field mouse | = | mouse in field |
| horsedoctor | = | doctor for horse |
| olive oil | = | oil from olive |
| tax law | = | law about tax |

Levi gives the traditional equivalents for her recoverably deletable predicates as given in the table below (Levi, 1978:77):

| RDP | Traditional term |
|---|---|
| CAUSE | causative |
| HAVE | possessive/dative |
| MAKE | productive, constitutive, compositional |
| USE | instrumental |
| BE | essive/appositional |
| IN | locative [spatial or temporal] |
| FOR | purposive/benefactive |
| FROM | source/ablative |
| ABOUT | topic |

Levi (1978:50-51) lists the following as her fundamental claims and basic principles of her theory:

1. Complex nominals are all derived from an underlying NP structure containing a head noun and a full S in either a relative clause or NP complement construction; on the surface, however the complex nominal is dominated by a node label of N.
2. Semantic restrictions are identified for complex nominals and for the propositions contained in their underlying structures, in the sense that they are no more and no less idiosyncratic for the former than for the latter.
3. Any given CN form is inherently and regularly ambiguous over a predictable and relatively limited set of possible readings; although any one of these readings may be used more frequently than the others, or even exclusively, in a given speech community and during a certain period, the potential ambiguity still remains part of a

speaker's competence and hence must be recognised in any grammatical description of these forms.

4. Complex nominals are all derived by just one of two syntactic processes: the deletion or the nominalization of the predicate in the underlying S.
5. For complex nominals derived by predicate deletion, a small set of Recoverably Deletable Predicates (RDPs) can be specified such that only its members, and no other predicates, may be deleted in the formation of CNs; the members of this set are CAUSE, HAVE, MAKE, BE, USE, FOR, IN, ABOUT, and FROM.
6. The potential ambiguity that is created by the multiplicity of possible underlying sources for a given surface CN is drastically reduced in discourse by both semantic and pragmatic considerations.
7. The relationship expressed by the RDPs appear to be of such semantic primitives that the set of RDPs proposed herein for English may well reflect universal constraints on the semantic structure of complex nominals in all languages; preliminary evidence from a number of other languages, as well as from other areas of the grammar, supports this hypothesis.
8. CN formation involves some process which manifest the generality of regular, nongoverned syntactic transformations, and others which have more in common with the idiosyncrasies that characterize the areas of both morphological derivation and lexical entries; thus, an account of CN formation will differ in certain respects form accounts of fully general syntactic process as well as from accounts of either partially productive derivational processes or totally idiosyncratic lexical entry rules.

Though we do not follow the transformational generative approach taken by Levis in the derivation of nominal compounds, certain insights given by her regarding the recoverability of the meaning of the compounds by positing a minimal set of underlying predicates is worth taken into consideration for the stand we have taken up here.

**Vijayavenugopal's approach**

Vijayavenugopal (1979), having influenced by the insights of Lees (1960, 1970), Botha (1968), Reibel (1963) and Fillmore (1968), as claimed by him, has taken up transformational generative grammarian's view of his time. According to him nominal compounds in Tamil are derived from their deep structure representations by transformation. For example the compound, *maNpommai* 'earthen-toy' is derived from the underlying sentence *oruvan pommaiyai maNNaal ceytaan* 'somebody made the toy in earth'. Accordingly he has classified the compounds into ten types out of which the 6th one is miscellaneous, 7th one consists of co-ordinate compounds, 8th one consists of figurative nominal compounds 9th one consists of idiomatic nominal compounds and 10th one consists of multinominal compounds. Here we are interested in his first six types of nominal compounds. As for as these compounds are concerned he takes the position that the

consituents of a compound are related to each other by certain relations (which include case relations) established by an underlying verb. In the first class of nominal compounds, the second member is a subject and the first member is any one of the following: subject, ablative, instrument, purposive, locative, temporal, adverb, and verb. In the second class of nominal compounds, the second member is a cognate object and the first member is a subject or object or instrumental. In the third class of nominal compounds the second member is subject and the first member is any one of the following: object, sociative, dative, ablative, locative, temporal, adverb, cause and predicate. In the fourth class of nominal compounds the second member is locative and the first member is subject or object. In the fifth class of nominal compounds the second member is temporal and the first member is subject or object. The sixth class of nominal compounds contains the following subtypes: specific + generic, object + verb + (subject) *-i*, locative + verb + (subject) *-i* + verb and object + verb + (instrument) *-i*. The following table will exemplify his classification:

| **Sr. No** | **Class & Subclass** | **Compound (example)** | **Expansion of compound** |
|---|---|---|---|
| 1.1 | Subject + Object | *kaakkak kuuTu* 'nest of crow' | *kaakkaa kuuTu kaTTiyatu* 'crow built nest' |
| 1.2 | Instrument + Object | *maN pommai* 'clay doll' | oruvan pommaiyai maNNaal ceytaan 'somebody made doll by clay' |
| 1.3 | Ablative + Object | *karumpuccaaRu* 'sugar-cane juice' | *oruvan karumpiliruntu caaRRai eTukkiRaan* 'someone extracts the juice from the sugar- cane' |
| 1.4 | Purposive + Object | *tuNippai* 'cloth bag' | *oruvan tuNivaittuk koLvataRkaaka piyai vaittirukkiRaan* 'someone is keeping a bag in order to keep cloth' |
| 1.5 | Locative + Object | *neRRiccuTTi* 'ornament (worn) on the fore-head' | *orutti neRRiyil cuTTi kaTTukiRaaL* 'someone (a woman) ties an ornament on the fore-head' |
| 1.6 | Temporal + Object | *matiya uNavu* 'noon meals' | *avarkaL matiyattil uNavaik koTukkiRaarkaL* 'they give meals in the noon' |
| 1.7 | Adverb + Object | *tappukkaNakku* 'wrong calculation' | *oruvan kaNakkut tappaakap pooTTaan/ceytaan* 'someone did the calculation/sum wrongly' |
| 1.8 | Verb (predicate) + Object | *kuTi taNNiir* 'drinking water' | *oruvan taNNiir kuTikkiRaan* 'someone drinks water' |
| 2.1 | Subject + Cognate object | *karaTip piTi* 'bear's hug' | *karaTi piTi piTittatu* 'the bear hugged' |

| | | | |
|---|---|---|---|
| 2.2 | Object + Cognate Object | *teeni vaLarttal* ‘bee keeping’ | *oruvan teeniyai vaLarkkiRaan* ‘someone rears the bees’ |
| 2.3 | Instrumental + Cognate Object | *uTal uzhaippu* ‘physical labour’ | *oruvan uTalaal uzhaikkiRaan* ‘someone is working with (his) body’ |
| 3.1 | Object + Subject | *paal maaTu* ‘milk-cow’ | *maaTu paal tarum* ‘cow gives milk’ |
| 3.2 | Sociative + Subject | *naaTTup paRRu* ‘patriotism’ | *paRRu naaTTooTu uNTu* ‘there is love towards the country’ |
| 3.3 | Dative + Subject | *teeL koTukku* ‘sting of a scorpion’ | *teeLukku koTukku uNTu* ‘there is sting for the scorpion’ |
| 3.4 | Ablative + Subject | *kuzhaayt taNNiir* ‘tap water’ | *taNNiir kuzhaayiliruntu varukiRatu* ‘water comes from the tap’ |
| 3.5 | Locative + Subject | *kaTal miin* ‘sea fish’ | *miin kaTalil vaazhkiRatu* ‘fish lives in sea’ |
| 3.6 | Temporal + Subject | *anti mantaarai* ‘a type of flower which blooms in the evening’ | *mantaarai antiyil puukkum* ‘*mantaarai* blooms in the evening’ |
| 3.7 | Adverb + Subject | *teLi niir* ‘clear water’ | *niir teLivaaka irukkiRatu* ‘the water is clear’ |
| 3.8 | Cause + Subject | *paci mayakkam* ‘fatigue due to hunger’ | *paci eeRpTTataal mayakkam vantatu* ‘the fatigue came due to hunger’ |
| 3.9 | Predicate + Subject | *aaTu catai* ‘calf muscle’ | *catai aaTum* ‘muscle will move’ |
| 4.1 | Subject + Locative | *paampu puRRu* ‘serpent-hole’ | *paampu puRRil tankukiRatu* ‘snake lives in hole’ |
| 4.2 | Objective + Locative | *miTTaaykkaTai* ‘sweetmeat shop’ | *oruvan kaTaiyil miTTaay viRkiRaan* ‘someone sells sweetmeat in shop’ |
| 5.1 | Subject + Temporal | *mazhaikkaalam* ‘rainy season’ | *mazhai oru kuRippiTTa kaalattil peyyum* ‘it will rain in a particular period’ |
| 5.2 | Object + Temporal | *caappaaTTu neeram* ‘lunch time’ | *oruvan oru kuRippiTTa neerattil caappiTuvaan* ‘someone will eat food at a particular time’ |
| 6.1 | Specific + Generic | *tamizhk kuTi* ‘Tamil clan’ | *inta kuTi tamizh aakum* ‘this clan is Tamil’ |
| 6.2 | Object + Verb + (Subject) -*i* | *uur cuRRi* ‘wanderer’ | *oruvan uuraic cuRRukiRaan* ‘someone goes around the village’ |
| 6.3 | Locative + Verb + (Subject) -*i* | *teru poRukki* ‘rag picker’ | *oruvan teruvil poRukku- kiRaan* ‘someone picks up in the street’ |
| 6.4 | Object + Verb + (Instrument) -*i* | *maN veTTi* ‘spade’ | *oruvan maNNai karuviyaal veTTinaan* ‘someone cuts the earth with an instrument’ |

It is difficult to accept the deep structure representations posited by Vijavenugopal on the ground that there could be many sentential deep structures from which each compound can be derived and that some of them are very unnatural. Moreover the derivation from the sentential structures leaves open certain problems with regard to the deep-structure verbs, tense in verbs, gender of subject, etc. His positing of verb in past tense form for some compounds and in present tense or future tense form for some other compounds is quwstionable. Even the interpretation of an element as subject or object poses problem. For example, *veTTi* of *maN veTTi* can be interpreted as the subject as it can also be derived from the sentence karuvi *maNNai veTTum* 'the instrument cuts the earth'. Though we do not take up the transformational generative approach adopted by him as a whole, we can take into account certain insights given by him for our interpretive approach.

**Knowledge Based Representation approach**

As we have mentioned already that the position taken up by us is one of interpretation rather than generation from deep structures and that the process of nominal compound formation can be simply denoted by the rule, N + N ---> N, the computer analysis we propose here aims to interpret the meaning of the compounds from the information available in the nouns participating in the compound formation. We can take relevant clues from traditional grammars, Levi (1978) and Venugopal (1979). The representations of nominal compounds to interpret their meanings by the above mentioned three approaches show us that there are relations existing between the concatenated nouns which are to be established through certain verbs or predicates. That means we have to look forward for a grammatical formalism which can help us to interpret or retrieve the unexpressed part linking the constituents of a compound from the information available in the constituents themselves.

A knowledge-based system will serve our purpose. "Knowledge-based system emphasise meaning. Instead of processing data as a string of bits, they represent the meaning of data in terms of the real world. They carry on conversations with people in ordinary language, they find important facts before they are requested, and they solve complex problems at expert level of performance." (Sowa, 1983: Preface). Artificial intelligence (AI) and cognitive science are the two fields which are devoted to knowledge-based systems. Cognitive science takes into its fold philosophy, linguistics and psychology allowing a strong influence from computer science. Artificial intelligence is the engineering part and it focuses on programming tools and techniques than to philosophical issues. The present analysis makes use of both cognitive science and artificial intelligence. Conceptual graphs which emphasise meaning are widely used in AI systems.

Broadly speaking conceptual graphs are logical forms that state relationships between entities, attributes, and events. A detailed study of the notion of conceptual structure adopted here is available in Sowa (1984). The conceptual relations are similar to case relations

established by case oriented grammars. The sentences *A man bites a dog* and *A cat is sitting on a mat* can be represented by conceptual graphs as follows:

[MAN]⟵(AGNT) ⟵[BITE] ⟶ (OBJ) ⟶ [DOG]

[CAT] ⟵ (AGNT) ⟵[SIT] ⟶ (LOC) ⟶[MAT]

The information regarding tense and modalities are omitted in the above graphs. The square brackets enclose concepts and the braces enclose conceptual relations. "(AGNT)", "(OBJ)" and "(LOC)" denote respectively the conceptual relations (similar to case relations) AGENT, OBJECT and LOCATION. Though the conceptual relations are finite in number, it is not attempted here to establish the whole set of conceptual relations for Tamil as it will take us to a different direction. The conceptual relations which are needed for the interpretation of compound nouns are already in vogue in case oriented grammars and are self-explanatory

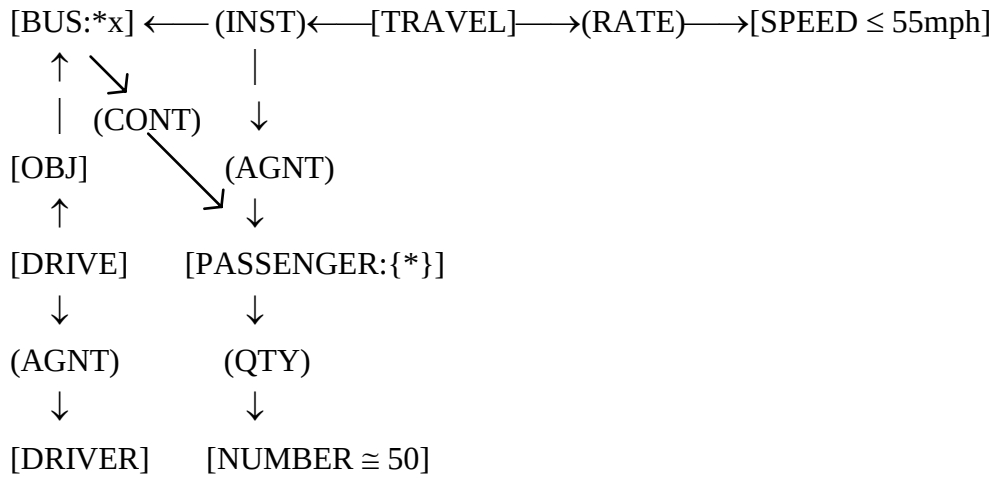
. There are at least three levels of complexity of conceptual graphs which we have to account here:

1. Arbitrary conceptual graphs which impose no constraints on permissible combinations.
2. Canonical graphs which enforce selectional constraints. They correspond to the case frames in linguistics and the category restrictions in philosophy.
3. Schemata which incorporate domain-specific knowledge about the typical constellations of entities, attributes, and events in the real world.

"A canonical graph is a combination of concept nodes and relation nodes where every arc of every conceptual relation is linked to concept." (Sowa, 1984:90). As not all such combinations make sense and some of them include absurd combinations like the following (which is an out come of Chomsky's famous example *Colourless green ideas sleep furiously.*) certain graphs are declared canonical:

[IDEA] ⟵(AGNT) ⟵[SLEEP] ⟶(MANN) ⟶[FURIOUSLY]
↓
(ATTR)
↓
[COLOURLESS GREEN]

Canonical graphs are meaningful graphs which represent real or possible situations in the external world. Schema is the basic structure for representing background knowledge for human like inference. Schemata favour plausible combinations by incorporating more knowledge about the world and avoiding less likely possibilities. While canonical graph represent everything that is conceivable, schemata represent everything that is possible. The following is the schema for BUS (Sowa,1984:129):




(In the above diagrammatic representation of schema for BUS, "*x" is a variable in which "*" is a generic marker and "x" is an indicator for cross reference, "(INST)" denotes the relation INSTRUMENT, "(CONT)" denotes the relation CONTAIN, "(QTY)" denotes QUANTITY, " ≤ " denotes 'lesser than or equal to' and " ≅" denotes range.)

**COMPUTATIONAL ANALYSIS OF NOMINAL COMPOUNDS**

Our proposal for automatic interpretation of meaning of the compounds from the meanings of concatenated nouns leads us to turn our eyes to an approach founded on cognitive science and artificial intelligence of knowledge representation. Conceptual graphs will be used to serve our purpose. From the meanings of constituent nouns (which constitute a compound) which are expressed in terms of conceptual graphs, it is possible for us to establish the meaning of the compound.

Take for example the compounds *nooy kirumi* 'germs which causes disease'. A schema of *kirumi* = GERMS should contain the information that 'germs can cause disease' and similarly a schema of *nooy* = DISEASE should contain the information that 'disease can be caused by germs'. That means the schemata for both the constituent nouns contain the following common graphical portion which might have been expressed differently :

[GERMS] ⟵(AGNT) ⟵[CAUSE] ⟶(OBJ) ⟶[DISEASE]

The unification of these two graphical portions can give us the following conceptual graph for the compound *nooy kirumi*:

[DISEASE CAUSING GERMS]⟵(AGNT)⟵[CAUSE]⟶(OBJ)⟶[DISEASE]

The schema for the *kaakkaay* = CROW and *kuuTu* = NEST should contain the information that 'crow can build nest' and 'nest can be built by crow' respectively. The following graphical portion will be shared by both the constituent nouns:

[NEST] ⟵ (OBJ) ⟵ [MAKE] ⟶ (AGNT) ⟶ [CROW ⊂ BIRD]

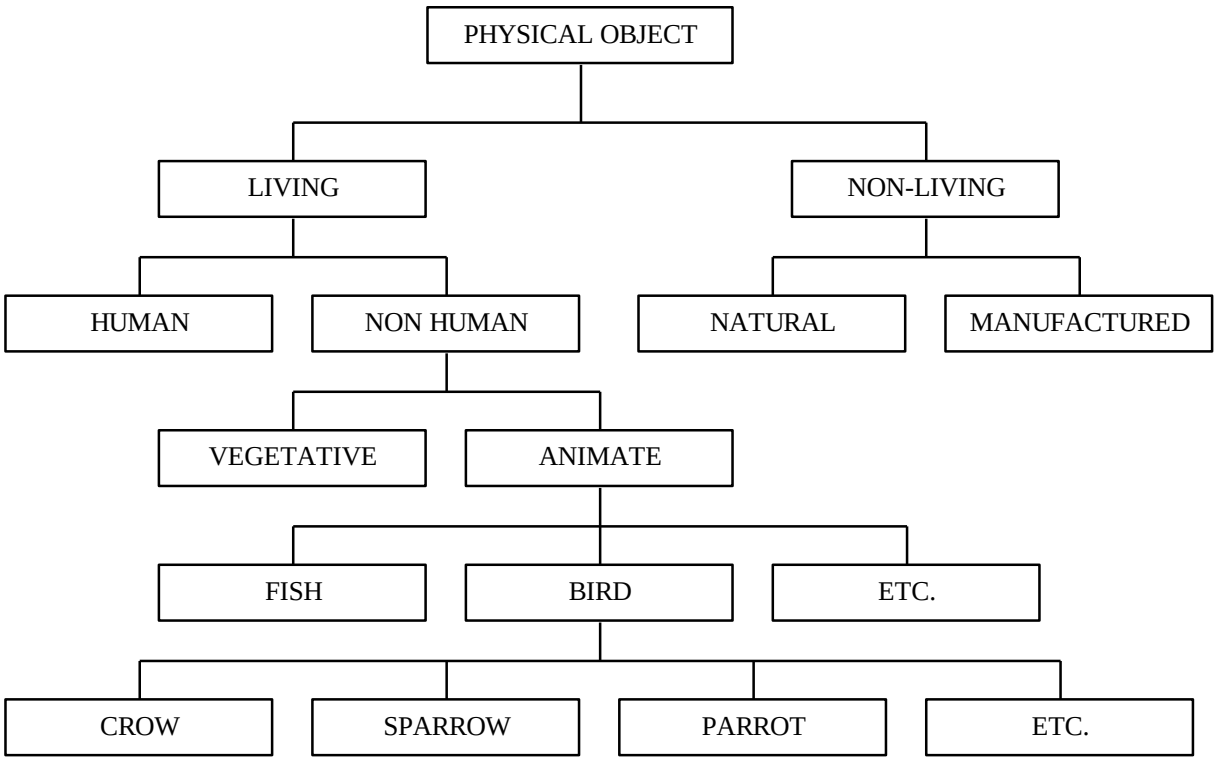
" ( ⊂ ) " should be read here as 'sub-set of')

The meaning of the compound *kaakkaaykkuuTu* can be interpreted by the unification of the relevant portions of the schemata. The schema for *maN* = CLAY as well as *pommai* = TOY should contain the information that 'clay can be used to make toy' and 'toy can be made up of clay' respectively. The following graphical portion will be shared by both the constituent nouns:

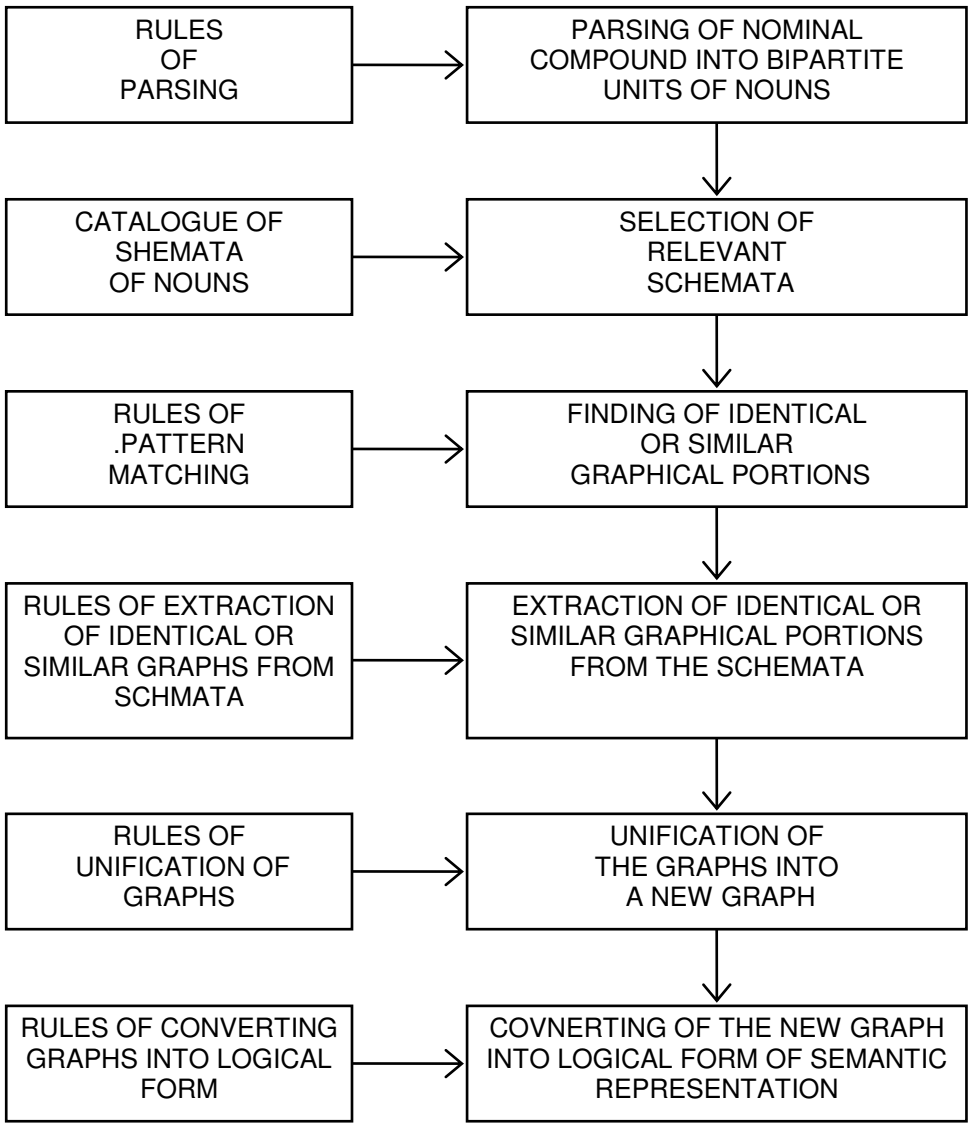
[TOY]←—(OBJ)←—[MAKE] —→(AGNT) —→[MAKER]
↓
(MADE UP OF)
↓
[CLAY ⊂ MATERIAL]

The meaning of the compound *maN pommai* can be interpreted by the unification of the relevant graphical portions of the schemata of the constituent nouns.

There is no need of cataloguing of schema for each concept as the schema of a concept at a higher level can be used to build the schema for a concept at lower level. For example the schema for BIRD can be used to build schema for CROW and *kuruvi* = SPARROW and also the interpretation of *paRavaikkuuTu* 'bird's nest' can take care of the interpretation of *kaakkaik kuuTu* 'crow's nest' and *kuruvikkuuTu* 'sparrow's nest'. The hierarchical representations of canonical graphs can be captured by having a thesaurus in which the concepts are arranged in the hierarchical fashion. The following is a tesauric model for PHYSICAL OBJECT:




The computational process of interpreting the meaning of nominal compounds from the meanings of the nouns which constitute them can be depicted by the flow chart given below:


RULES OF PARSING
PARSING OF NOMINAL COMPOUND INTO BIPARTITE UNITS OF NOUNS
CATALOGUE OF SHEMATA OF NOUNS
SELECTION OF RELEVANT SCHEMATA
RULES OF .PATTERN MATCHING
FINDING OF IDENTICAL OR SIMILAR GRAPHICAL PORTIONS
RULES OF EXTRACTION OF IDENTICAL OR SIMILAR GRAPHS FROM SCHMATA
EXTRACTION OF IDENTICAL OR SIMILAR GRAPHICAL PORTIONS FROM THE SCHEMATA
RULES OF UNIFICATION OF GRAPHS
UNIFICATION OF THE GRAPHS INTO A NEW GRAPH
RULES OF CONVERTING GRAPHS INTO LOGICAL FORM
COVNERTING OF THE NEW GRAPH INTO LOGICAL FORM OF SEMANTIC REPRESENTATION

Algorithm:

1. Parse the nominal compounds into bipartite units of constituent nouns.
2. Select the schema for each noun which constitute the compound from the conceptual catalogue of schemata.
3. Identify the similar or identical graphical portions of schemata.
4. Extract the identical portions.
5. Unify them in such a way that the head-nounship is not altered.
6. Covert the conceptual graph into logical proposition which represent the meaning of the compound.

## CONCLUSION

The proposed alternative of automatic interpretation of the meaning of nominal compounds from the meanings of nouns which constitute them is aimed to recapitulate the psychological reality of understanding by human brain in which the percepts are stored as conceptual graphs and the building up of and understanding of the new complex graphs is based on the already existing or stored conceptual graphs. Though knowledge representation by conceptual graphs is a tedious process and basing the semantic interpretation of compounded lexical units on the graphical representations of concepts is a challenge and sometimes may appear as a futile task, such attempt will depict the psychological reality of understanding by means of the information stored in the brain in terms of perceptual models called conceptual graphs. The deep structure representations from which the compounds are derived by transformational rules or formative rules is artificial and analysing and understanding a derived language unit by making use of a formalism whose foundation is deep structure representation does not represent a model which tries to capture the cognitive process of understanding and generation of language units. As conceptual graphs are finite they can be easily stored and manipulated by the computer.

The conceptual semantic interpretation of nominal compounds helps in the translation of the compounds into other language. This type of conceptual graphic representation lend a helping hand in the machine translation in which the interlingual representation of compounds is a challenging job.

## REFERENCE


Allen, M. 1978. Morphological investigation. PhD dissertation, University of Connecticut.

Aronoff, M. 1976. Word formation in generative grammar. Cambridge, MA: MIT Press.

Bauer, L. 1983. English word-formation. Cambridge: Cambridge University press.

Botha, R.P.1968. The function of lexicon in transformational generative grammar.

Fabb, N. 1984. Syntactic affixation. PhD. dissertation, MIT.

Beard, R. 1988. On the separation of derivation from morphology: Towards a lexeme/morpheme-based morphology. Quaderni di Semantica 9, 3-59.

Borer, H. 1988. On the parallelism between compounds and constructs. Year book of morphology 1, 45-66.

Botha, R.P. 1968. The Function of lexicon in transformational generative grammar. The Hague: Mouton.

Botha, R.P. 1983. Morphological mechanism. Oxford: Pergamon press.

Chomsky, N. 1970. Remarks of nominalization. In: Jacobs, R and Rosenbaum, P. (eds.). Readings in English transformational grammar. Waltham, MA: Blaisdell.

Di Sciullo, A.M. and Williams, E. 1987. On the definition of word. Cambridge, MA: MIT press.

Fillmore, C.F. 1968. Case for case. In: Batch, E. and Harms , R. (eds.). 1968. Universal in Linguistic theory. New York: Holt, Rinehart and Wilson.

Lees, R.B. 1960. The Grammar of English nominalizations. International Journal of American Linguistics 26.

Lees, R.B. 1963. The Grammar of English nominalization. The Hague: Mouton.

Lees, R.B. 1970. Problems in the grammatical analysis of English nominal compounds. In: Bierwisch, M. and Heidolph, K.L. (eds.). 1970. Progress in Linguistics. The Hague: Mouton.

Levi, J.N. 1978. The Syntax and semantics of complex nominals. New York: Academic press.

Lieber, R. 1983. Argument linking and compounding in English. Linguistic Inquiry, 14, 251-86.

Roeper, T. 1987. Implicit arguments and head-complement relation. Linguistic Inquiry 18, 267-310.

Roeper, T. 1988. Compound syntax and head movement. Year book of morphology. 1, 187-228.

Roeper, T. and Siegel, D. 1978. A Lexical transformation for verbal compounds. Linguistic Inquiry 9, 199-260.

Selkirk, E. 1982. The Syntax of words. Cambridge, MA: MIT press.

Sowa, J.F. 1984. Conceptual structures: Information processing in mind and machine. Reading: Addison-Wesley publishing company.

Spencer, A. 1991. Morphological theory: An Introduction to word structure in generative grammar. Oxford: Basil Blackwell.

Sproat, R. 1985. On deriving the lexicon. PhD dissertation. MIT.

Vijayavenugopal, G. 1979. Nominal composition in Tamil. Madurai: Madurai Kamaraj University.

**Note: The paper was published in South Asian Language Review, vol. 7, no. 1, 1997, 27-46.**